8-11-2020

Delusions as Hetero-Dynamic Property Clusters

Shelby Clipp

DELUSIONS AS HETERO-DYNAMIC PROPERTY CLUSTERS

by

SHELBY CLIPP

Under the Direction of Neil Van Leeuwen, PhD

ABSTRACT

The standard position in psychiatry maintains that delusions are *beliefs*. However, the features of delusions often diverge from those typically associated with belief. This discrepancy has given rise to what I refer to as the *doxastic status debate,* which concerns whether delusions are best characterized as "beliefs." Despite efforts, there has been little progress in settling this debate. I argue that the debate has been stymied because it's largely a verbal dispute (Chalmers, 2011). I then attempt to advance the debate into substantive territory by putting forward the *hetero-dynamic property cluster* (HDPC) model. According to the HDPC model, delusions are mental states characterized by an odd and unstable cluster of features, which, when manifested, gives delusions the appearance of straddling the line between two distinct types of attitudes.

INDEX WORDS: Delusions, Cognitive attitudes, Verbal dispute, Doxasticism, Property clusters

DELUSIONS AS HETERO-DYNAMIC PROPERTY CLUSTERS

by

SHELBY CLIPP

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2020

DELUSIONS AS HETERO-DYNAMIC PROPERTY CLUSTERS

by

SHELBY CLIPP

Committee Chair: Neil Van Leeuwen

Committee: Dan Weiskopf

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

August 2020

ACKNOWLEDGEMENTS

I wish to thank Neil Van Leeuwen, for his guidance, support, and insightful discussions. I also wish to thank Dan Weiskopf, who provided many carefully considered comments on this paper. Finally, I wish to express my gratitude to Scott Buhrman, my high school teacher who not only introduced me to philosophy but also to the possibility of pursuing higher education; Tim Jankowiak, my undergraduate professor whose mentorship has been invaluable to my trajectory in academic philosophy, and; P.B.F., who not only helped me think about this project by listening to my ideas, but also helped me maintain my sanity along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

Ms. A, a 45-year-old woman, was brought to the emergency room after assaulting her husband. During her visit to the emergency room, Ms. A reported that her husband was not really her husband, that her two teenage sons had been murdered, and that the two boys living with her were strangers merely pretending to be her actual sons (Sharma et al., 2009). Ms. A's assertions are emblematic of Capgras delusion, the delusion that someone close to you such as a spouse or family member has been replaced by an identical impostor. Other bizarre assertions often indicative of a clinical delusion include: "Another person's thoughts are being inserted into my mind," "I am dead," "I'm being followed by people I know in disguise," "People are out to get me," "The radio is speaking directly to me."[1]

Delusions are peculiar mental states whose nature is controversial. Identifying which cognitive attitude best characterizes delusions (e.g., whether delusions are beliefs or another type of cognitive attitude) has been regarded as a crucial step to understanding the nature of delusions. The standard position in psychiatry maintains that delusions are *beliefs*, albeit irrational beliefs.[2] In recent years, however, the view that delusions are best characterized as "beliefs" has come under scrutiny due to the way in which the features frequently exhibited in cases of delusions often noticeably diverge from those typically associated with beliefs (e.g., delusions tend to be highly resistant to counterevidence, they exhibit limited behavioral guidance, and they tend to be incoherent). As a result of this discrepancy, an extensive debate has emerged in the philosophy of psychiatry concerning whether delusions are best characterized as "beliefs" (see Bortolotti

[1] These assertions correspond to thought insertion, Cotard delusion, Frégoli delusion, persecutory delusion, and delusion of reference.
[2] Irrational because unlike "rational" beliefs, delusions tend to arise from little or no justificatory evidence at all.

2009, Currie & Ravenscroft 2002, Bayne & Pacherie 2005, Dub 2017 and, Miyazono 2019). I refer to this debate as the *doxastic status debate*.[3]

Despite decades of attention, little progress has been made in settling the doxastic status debate. In §3, I argue that the debate has been stymied because it's largely a verbal dispute (Chalmers, 2011), marked by an underlying agreement about the features of delusions and an insubstantial disagreement over whether delusions should be labeled "beliefs." Following this, in §4, I attempt to advance the debate into more substantive territory by putting forward a new descriptive model for characterizing delusional attitudes, which I refer to as the *hetero-dynamic property clusters* (HDPCs) model. According to the HDPC model, delusions are best understood as mental states characterized by an odd and unstable cluster of features, which, when manifested, contributes to delusion's failure to resolve neatly into any one kind of pre-theoretical cognitive attitude. Throughout §4, I attempt to show how the HDPC model is clearer and more precise than insistence on characterizing delusions with pre-theoretical attitude terms such as "beliefs" and "imaginings." I begin with a general discussion of the doxastic status debate (§2).

2 THE DOXASTIC STATUS DEBATE

Cognitive attitudes refer to mental states that represent one's stance (i.e., attitude) towards some content.[4] Cognitive attitudes (such as *to believe*, *to think*, and *to imagine*) contrast with conative attitudes in that the former aim to represent how the world or some state of affairs truly is or how it might be, whereas conative attitudes (such as *to desire, to hope,* and *to wish*) express how the world or some state of affairs may be made to be.

[3] I occasionally refer to the doxastic status debate as simply "the debate."
[4] Attitudes can be expressed in the form "*S* A that *c,*" where *S* refers to the individual possessing the mental state, A refers to the attitude that *S* holds, and *c* refers to some content. For example, Ada *believes* that the frisbee is on the hill.

Ultimately, the doxastic status debate is a debate over which cognitive attitude best characterizes delusional mental states. Within the debate, scholars fall into one of two camps: doxastic and non-doxastic. Doxastic accounts, which have dominated the psychiatric and philosophical literature, maintain that delusions are beliefs. Indeed, the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) defines delusions as:

> A false belief based on incorrect inference about external reality that is firmly sustained despite what almost everybody else believes and despite what constitutes incontrovertible and obvious proof of evidence to the contrary" (American Psychiatric Association 2013, p.819).[5]

Unlike with doxastic accounts, there isn't a consensus among non-doxastic theorists as to which cognitive attitude best characterizes delusions. Rather, what unifies non-doxastic accounts is the view that delusions are *not* beliefs. For instance, Currie and Ravenscroft (2002) argue that delusions are peculiar *imaginings* misidentified by the individual as beliefs, whereas Dub (2017) argues that delusions are *acceptances* produced by pathologically powerful cognitive feelings.

In general, debates over which cognitive attitude best characterize a particular type of mental state are effectively disagreements over the functional roles played by the mental state in question.[6] Below, I discuss the functional roles of two of the most common cognitive attitudes proposed in the doxastic status debate, viz., beliefs and imaginings.[7] An important caveat to what follows is that the features presented are characteristic of *prototypical* beliefs and imaginings:

---

[5] It's worth noting that this definition is a contentious one. For example, Coltheart (2007, p.1043) asks whether a true belief could be a delusion, just as long as the believer has no good reason for holding the belief? Similarly, Coltheart asks whether all delusions must be based on inference and whether delusions need to be about external reality (e.g., Cotard delusion, the delusion that one's body parts are dead, dying, or non-existent doesn't appear to be about external reality). Of course, Coltheart also asks whether it needs to be the case that delusions are beliefs, rather than say, imaginings.

[6] Although less frequently the case, a disagreement over which cognitive attitude best describes a particular type of mental state can be a disagreement over the etiology of the mental state in question. An account of this style will be discussed in §3.2.

[7] When I use the terms "beliefs," and "imaginings" (as well as any other standard type of cognitive attitude), I'm using the terms as they are traditionally used in the debate. The use of these terms will later be rejected due to their problematic nature within the doxastic status debate.

they express the normative functional roles of beliefs and imaginings. However, actual instances of beliefs and imaginings often fall short of satisfying each of their characteristic features. This caveat will be addressed further in §4.3.

2.1 Beliefs

In the philosophy of mind, the use of the term "belief" generally refers to the attitude one has whenever they take it to be the case that something is true. In other words, if *S* believes that *p*, then *S* takes *p* to be true. As will become apparent, the concern with truth is indispensable to the nature of belief. Indeed, it's the concern with truth which warrants the *responsiveness to evidence* of beliefs. Beliefs are responsive to evidence in two important ways. First, there should be sufficient reason(s) or evidence for why one comes to hold a particular belief. Second, in cases where strong contradictory evidence to one's belief becomes available, the individual should either appropriately revise their initial belief or abandon it altogether. For instance, suppose that Elaine has the belief that her roommate is at home, and that she believes this because it's 5pm and her roommate is almost always at home at this time. However, when Elaine searches for her roommate she cannot find her in any areas of the house. Because beliefs tend to be responsive to evidence, Elaine's inability to find her roommate should lead her to revise her initial belief that her roommate is at home.

Suppose however, that despite not being able to find her roommate, Elaine asserts that her roommate is at home. In such a scenario, we might be tempted to think that there's something wrong with Elaine or that she's lying (or wishing, or hoping, etc.) that her roommate is at home. We'd likely be less tempted to say that in this scenario, Elaine *believes* that her

roommate is at home.[8] The reason why we'd likely be reluctant to ascribe belief to Elaine in this scenario has to do with belief's concern for truth. In particular, if we grant that beliefs aim at truth and we know that Elaine has overwhelming contrary evidence to the belief that her roommate is home, then characterizing Elaine's attitude as a belief would suggest that she's capable of successfully ignoring contrary evidence whenever she sees fit. The issue here is that if one can willingly ignore strong counterevidence in order to maintain one of their current beliefs, then belief's concern for truth is greatly diminished. Hence, in order to maintain that beliefs are attitudes that one has whenever they take it to be the case that something is true, beliefs are typically characterized as being *involuntary.*[9]

Additionally, the standard position among philosophers of mind is that beliefs work in conjunction with desires to produce actions that will make the content of a desire true if the content of the belief is true. This is to say that beliefs exhibit *behavioral guidance*. For instance, if Elaine has the desire to discuss the upcoming rent with her roommate, but she believes that her roommate isn't currently at home, this belief-desire combination should bestow Elaine with a range of potential practical behaviors (calling her roommate, waiting for her roommate to return, etc.). Relatedly, another feature typically attributed to beliefs is *coherence:* one's beliefs should logically cohere with their other beliefs and intentional states. Ideally, coherence among one's beliefs helps individuals reject any behaviors (or other beliefs) that would conflict with their

[8] Rose et al. (2017) argue, however, that behavioral circumscription does not undermine folk belief ascription, and that the type of evidence required for belief ascription is *lexicographically ordered* (2017, p.8). This is to say that when laypeople are given the choice to ascribe belief to an individual, it's the individual's sincere verbal assertions that tips the scale in favor of belief ascription, even when behavioral evidence contradicts the individual's assertion. However, an issue with Rose et al.'s conclusion on the nature of folk belief ascription is that the measures used in their study oversimplify folk belief ascription by appealing to a highly dichotomous view of belief. Furthermore, Rose et al. (2014, 2017) primarily focus on how behavioral circumscription influences folk belief ascription, and whether or not unresponsiveness to evidence, a hallmark of delusions, undermines belief ascription in cases where one sincerely asserts a proposition remains an open question.

[9] Not every philosopher is convinced that beliefs are involuntary, however. See Bennett (1990) and Ginet (2001) for more on this topic.

other beliefs about the world. For example, if Elaine's beliefs are coherent, she shouldn't, for instance, carry on a one-sided conversation about the rent while standing in her roommate's empty room since her other beliefs about the world should make it clear to her that an action of this sort would be illogical and unproductive.

A final feature of belief that has received less attention than the others is *context independence.* Context independence will be discussed in greater detail in the final section of this paper. Presently, the notion of context independence can be expressed by stating that if one believes *p*, *p* is taken as a given in practical reasoning and decisions, regardless of the practical setting one is in: where practical setting is "the dimension of a cognitive attitude (independent or dependent) that specifies the sorts of situations in which it guides behavior" (Van Leeuwen, 2014, p.702).

Although the features of belief remain a lively topic of debate, philosophers generally associate the features discussed above (*responsiveness to evidence, involuntariness, behavioral guidance, coherence*, and *context independence*) as characteristic of belief. Importantly, and what's hopefully apparent from Elaine's scenario, is that the features of belief tend to reinforce one another (i.e., the set of features are likely to co-occur with one another). This idea of the features of belief reinforcing one another will play a significant role throughout §4.

2.2 Imaginings

Although imaginings have received less philosophical attention than beliefs, the literature suggests that the features typically associated with imaginings are almost directly opposed to those associated with beliefs (see Currie, 2000; 2002, Nichols, 2004, Van Leeuwen, 2011). Furthermore, just as the features of belief tend to reinforce one another, so too, do the features of imaginings. Crucially, unlike with beliefs, concern for truth doesn't play an essential role in

imaginings. As a result, imaginings are typically *unresponsive to evidence*. That is, given that belief's responsiveness to evidence is bolstered by a concern for truth, the pressure for imaginings to be responsive to evidence is largely absolved. Moreover, by not aiming to reflect some state of affairs as actually being a certain way (and therefore, by not being required to be responsive to evidence), there's nothing which prevents imaginings from being *voluntary*. Perhaps unsurprisingly, imagining's lack of concern with truth entails that they *needn't be logically coherent*.

Until this point, the features typically associated with imaginings appear to be almost directly opposed to those associated with beliefs. However, both beliefs and imaginings can (and do) exhibit behavioral guidance. For instance, a child who imagines that the living room floor is covered in lava might hop around from one piece of furniture to the next in order to avoid touching the floor. Yet, when compared to beliefs, imaginings tend to exhibit *limited behavioral guidance* insofar as they tend to only guide behavior within the context of make-believe and not outside of this context. In other words, imaginings are *context dependent.* Table 1 below compares the features of beliefs and imaginings. In §4.2, this chart will be expanded to include the features of delusions.

Table 1: Attitude Features: Belief and Imagining

Attitude	**Belief**	**Imagining**
Feature		
Responsiveness to Evidence	**Yes**	**No**
Voluntary	**No**	**Yes**
Context Independent	**Yes**	**No**
Behavioral Guidance	**Yes**	**Yes (limited)**
Coherence	**Yes**	**No**

2.3 Characterizing Delusions

Because debates over which cognitive attitude best characterize a particular type of mental state are effectively disagreements over the functional roles played by that mental state, we should expect doxastic theorists and non-doxastic theorists to posit different functional roles of delusions. Moreover, provided this understanding of debates over attitude characterization, we should expect non-doxastic theorists who propose different attitudes from one another to also put forward different functional roles of delusions. Indeed, the very existence of the doxastic status debate suggests that doxastic and non-doxastic theorists take themselves to have substantively different views on delusions. In the next section, however, I argue that there's not a substantial difference between doxastic and non-doxastic accounts of delusions because the two parties are largely engaging in a *verbal dispute.*

3 VERBAL DISPUTES

A dispute typically arises when two parties disagree over the truth value of some claim. Sometimes the word choice in a dispute is significant. Other times, however, two disputing parties take themselves to have substantively different views about some domain when there's actually little more at stake than whether a certain term applies to the domain in question. In such cases, the disagreement among the two disputing parties is often merely terminological. That is, the disagreement is over insubstantial linguistic matters on which nothing about the domain in question depends. Such disputes have been referred to as *verbal disputes* (Chalmers, 2011).

The hallmark of a verbal dispute is when two parties agree on all of the *relevant* facts about a domain in question yet disagree on whether a certain label or term applies to a particular object of that domain. For instance, suppose two individuals, George and Elaine, find themselves

in a disagreement with one another. George insists that sandals are shoes and Elaine insists that sandals are not shoes. When George and Elaine defend their respective views, it's apparent that they both agree on all of the properties of sandals. Yet, according to Elaine, sandals aren't shoes because the term "sandal" refers to open-toed footwear whereas "shoe" refers to closed-toed footwear.

It's possible that George and Elaine are engaging in a verbal dispute insofar as their disagreement isn't over the properties of sandals, but merely over whether or not the sortal "shoe" applies to open-toed footwear. However, a disagreement over the use of a particular term isn't by itself sufficient for a verbal dispute. For it may be the case that there *is* something substantial about the use of a particular term, such as when the word choice in a dispute has significant practical ramifications. For instance, if George tells Elaine that wearing shoes inside of Kim's home is disrespectful, Elaine risks offending Kim if she insists on wearing sandals inside of Kim's home. The upshot here is that whether a dispute over the truth of a particular sentence is verbal doesn't depend on the sentence alone; it also depends on the background of agreement and disagreement of the parties in the dispute.

> The same sentence *S* can typically be the focus of both verbal and substantive disputes, depending on this background. For example, against the background of agreement on the motion of a squirrel, a dispute over "The man goes around the squirrel" may be verbal, but against a different background (perhaps we agree that the man goes around some animal but disagree about whether it is a squirrel or a mouse), the dispute may be substantive (Chalmers, p.518).

Evidently, it's not always obvious whether a dispute is a verbal dispute. However, a useful strategy for helping to determine whether a dispute over the truth of a sentence, *S*, that uses a term, *T,* is verbal in nature is Chalmers's *method of elimination*:

> First: one bars the use of term *T.* Second: one tries to find a sentence *S'* in the newly restricted vocabulary such that the parties disagree nonverbally over *S'* and such that the disagreement over *S'* is part of the dispute over *S*. Third: if there is such an *S',* the dispute

> over *S* is not wholly verbal, or at least there is a substantive dispute in the vicinity. If there is no such *S'*, then the dispute over *S* is wholly verbal (Chalmers, p.526-7).

Applying the method of elimination to the debate then, we can ask whether doxastic and non-doxastic parties can state what they're disagreeing about without using the term "belief." But prior to applying the method of elimination to the debate, we must first identify the parties' background of agreement and disagreement.

3.1 Verbal Disputes and the Doxastic Status Debate

To understand the parties' background of agreement and disagreement, the first thing to note is that by and large both doxastic and non-doxastic theorists agree on the prototypical features of belief (§2.1), and that prototypical imaginings tend to lack these features (§2.2). Strikingly, the two parties *also* tend to agree on the specific features of delusions and on how these features diverge from those of prototypical beliefs and imaginings. In fact, contrary to what we might expect from a substantive dispute over cognitive attitudes, a detailed inspection of the doxastic status debate literature reveals that both doxastic and non-doxastic theorists largely agree on the following features as being characteristic of delusions: (i) involuntariness; (ii) an unusual degree of unresponsiveness to evidence and insensitivity to evidence (delusions tend to arise from either no evidence or from "evidence" that seems patently unfit to justify the content of the delusion); (iii) lack of behavioral guidance (i.e., behavioral circumscription), and; (iv) lack of coherence.[10]

For example, in her influential doxastic account, Lisa Bortolotti (2009) maintains that delusions are unresponsive to evidence.[11] According to Bortolotti, however, unresponsiveness to

[10] See Miyazono (2019), Dub (2017), Bortolotti (2009), Egan (2009), Currie (2000) and Campbell (2001).
[11] Bortolotti states that most delusions are "epistemically irrational" in that they're unsupported by available evidence, as well as unresponsive to counterevidence and counterargument. According to Bortolotti, delusions are beliefs in the same way that non-delusional irrational beliefs (e.g. racist, religious, and superstitious beliefs) are:

evidence isn't sufficient for rejecting doxasticism since many "typical beliefs" also exhibit this feature:

> [N]ormal subjects also present biases in the gathering and weighting of evidence, and in the way in which they insulate certain beliefs from potentially disconfirming evidence. For instance, we all tend to be too conservative when asked to re-assess strongly held beliefs, and we are unlikely to modify such beliefs, even if convincing evidence against them becomes available (Bortolotti, p.116).

In agreement with Bortolotti, the non-doxastic theorist, Richard Dub (2017), also maintains that delusions are unresponsive to evidence. Yet, Dub reasons that delusions are unresponsive to evidence because they're not beliefs at all but *acceptances* which unlike beliefs, "can be, but are not necessarily, volitionally formed for prudential, and non-evidence-based reasons" (Dub, p.10). However, aside from characterizing delusions as "acceptances" rather than "beliefs," the features Dub attributes to delusions fail to differ in any obvious way from those put forward by doxastic theorists, such as Bortolotti.

We see a similar agreement among doxastic and non-doxastic theorists concerning the behavioral guidance of delusions. For instance, Greg Currie (2000), a non-doxastic proponent who proposes that delusions are peculiar imaginings, maintains that delusions lack belief-like behavioral guidance (i.e., delusions exhibit *behavioral circumscription*):

> [A delusional thought] exerts a powerful psychological force, absorbing inner mental resources, but it fails to engage behaviour in the way that genuine belief would (Currie, p.175).

By way of example, Currie mentions a schizophrenic patient who expresses paranoid ideas about her sister, but who never acts on her delusion (p.175).[12] Related to the behavioral circumscription

---

while both of these types of mental states may fail to exhibit some of the prototypical functional features of belief, their failure to do so merely indicates a difference in degree and not in kind. For example, delusions are *more* insensitive to evidence than non-delusional irrational beliefs, and non-delusional irrational beliefs are *more* insensitive to evidence than "normal" functioning beliefs.

[12] For instance, the patient neither harms nor avoid her sister.

of delusions, Currie's example also appeals to the phenomenon of *double-bookkeeping*, wherein delusional individuals appear to perceive both reality and their delusion as being real yet remain largely unaffected by the discrepancies between the two "realities."[13]

Once again, in line with Currie's non-doxastic account, Bortolotti (2011) maintains that delusions lack behavioral guidance (i.e., delusions exhibit behavioral circumscription) *and* that delusional individuals may experience double-bookkeeping. Bortolotti argues however, that the occurrence of double-bookkeeping and behavioral circumscription doesn't entail non-doxasticism because these features can be explained by a delusional individual's failure to sustain motivation to act on their beliefs (Bortolotti, 2011). Furthermore, although Currie and Dub are both non-doxastic theorists, they differ in terms of which attitude they propose best characterizes delusions. Yet, just as the doxastic and non-doxastic theorists put forward the same features of delusions, so too, do Currie and Dub.

Given the doxastic and non-doxastic parties' mutual agreement over the features of delusions, we should be left wondering what exactly the two parties are disagreeing over. If we were to apply Chalmers's method of elimination to the debate by barring the term "belief," it seems that we would be hard pressed to locate any substantive difference among doxastic and non-doxastic accounts. Instead, what the two parties actually appear to be disagreeing over whether the term "belief" should be used for an attitude that shares *x* features, but not *y* features, with "typical beliefs." Evidence that the doxastic status debate is largely a verbal dispute becomes even more compelling if we recognize that the peculiar features of delusions which

[13] An example of double-bookkeeping from Dub (p.33) comes from a report from an individual with schizophrenia who claimed to feel certain that time and space don't exist yet understood that she must walk down the street and get on the train to get to her psychiatry appointment. Although double-bookkeeping is perhaps most apparent in cases of schizophrenic delusions, it can manifest in all types of delusions including Capgras such as when one asserts that their spouse is an impostor but remains otherwise unphased about their delusion.

motivate non-doxastic proponents to argue that delusions should *not* be characterized as beliefs, are the very same features that doxastic proponents argue are not unusual enough to reject delusions as being beliefs.

Nevertheless, for reasons addressed at the beginning of this section, one could argue that although the parties' agreement over the functional roles of delusions is emblematic of a verbal dispute, it still remains possible for the debate to comprise a substantial theoretical disagreement whose primary focus is not the attitude *features* of delusions. Kengo Miyazono (2019) takes an approach of this style by presenting a doxastic teleo-functionalist account of delusions. I consider Miyazono's account in the following subsection. After addressing Miyazono's position, I speculate as to why scholars of the debate have been tempted to engage in a verbal dispute in the first place.

3.2 Miyazono's Malfunctioning Belief Hypothesis

> The category of heart…, is defined in terms of the distinctively heart-like function, i.e., the function of pumping blood. All (and only) members of this category *have* the function of pumping blood. But this does not mean that all the members of this category actually *perform* the function of pumping blood. Diseased or malformed hearts *have* the function of pumping blood and thus belong to the category of the heart, but they do not *perform* the function (Miyazono, p.4).

According to Miyazono (2019), delusions are *malfunctioning beliefs.* Miyazono argues that delusions are best characterized as beliefs because they have doxastic functions, yet similar to diseased or malformed hearts, delusions fail to perform the functions associated with functional beliefs due to malfunctioning cognitive mechanism(s). The structure of Miyazono's argument can be summarized as follows:

P1: Teleo-functionalism individuates mental states by their *function* (rather than by the mental states current or actual causal roles).

P2: Within a teleo-functionalist framework, something's having a certain function is a historical property.

C1: Thus, within a teleo-functionalist framework, if something has the right kind of evolutionary history, then it's possible for it to have a certain function without actually performing that function.

Integral to Miyazono's argument is that delusions arise from *belief* forming mechanisms. With this in mind, it seems that Miyazono has introduced a substantive disagreement into the doxastic status debate, viz., the etiological and causal source of the delusions. If his view is correct, then there does appear to be significant consequences of labeling delusions one way rather than another. Specifically, if the etiological and causal source of delusions is the appropriate focus of the doxastic status debate, and if delusions arise from *belief* forming mechanisms, then how we characterize delusions is significant in that gaining a better understanding of beliefs (and *not* of imaginings, acceptance, etc.) and their mechanisms should subsequently contribute to a better understanding of delusions.

Unfortunately, although Miyazono's teleo-functionalist account may be useful for characterizing objects such as hearts, it's debatable whether his account can be equally useful in the doxastic status debate. In particular, a potential challenge to Miyazono's account is whether it can actually handle a key difference between the attitudes and objects such as hearts. For in the latter case, we have available to us additional criteria for identifying the object in question, aside from the object's functions (e.g., physical appearance, location). Yet, in the case of attitudes, the criteria for identification available to us is primarily the attitude's functions (whether the attitude

is unresponsive to evidence, whether it's involuntary, and so forth). Hence, our ability to identify malfunctioning attitudes appears to be much more difficult than identifying objects such as malfunctioning hearts. Moreover, even if we were to empirically identify belief forming mechanism(s) in a manner similar to our empirical identification of hearts, it seems unlikely that these mechanisms (theory of mind mechanisms, perception mechanisms, etc.) would be distinct from the mechanisms involved in other cognitive attitudes, including imaginings (see Nichols, 2004). Thus, to say that delusions are the result of malfunctioning *belief* mechanisms comes across as a bit of a stretch.

So, while I'm sympathetic to Miyazono's position, and I believe it's likely that delusions involve malfunctioning cognitive mechanisms of some sort, I'm not convinced that Miyazono's account supports the view that delusions are malfunctioning *beliefs.* If Miyazono's doxastic account is to succeed, there are at least a few things in need of further support. Such as the ability to distinguish between belief and other cognitive attitudes at the cognitive mechanism level; the ability to flesh out an account of doxastic functions that's exclusive to belief or an account which can non-arbitrarily draw the line between the functions of beliefs and belief-like cognitive attitudes; and perhaps most importantly, the ability to show that these doxastic functions are what's malfunctioning in cases of delusions rather than some other non-doxastic functions or cognitive processes.

3.3 Tacit Essentialism

If the doxastic status debate is in fact a verbal dispute, then the current direction of the debate presents a major stumbling block to our understanding of the nature of delusions. In an effort to steer the debate onto a more productive path, there's value in attempting to pinpoint what exactly has led the scholars of the debate to be tempted by the verbal dispute in the first

place. To be fair, uncovering the full picture of what has led scholars to engage in a verbal dispute is likely a complex issue with many contributing factors. That said, below I speculate that the verbal dispute is at least partially related to the way in which philosophers tend to think about attitudes and natural kinds.

The ability to organize objects into categories plays an important role in human cognition (Foster-Hanson, et al., 2016). In a now influential article, Rosch et al. (1976) state that "one of our most basic functions is the act by which we separate the environment into classifications by which nonidentical stimuli can be treated as equivalent" (p.382); where, classification refers to the construction of categories (generally designated by names, e.g., banana, fruit) or groups of entities on the basis of their "important"[14] shared attributes or relations. In philosophical literature, natural kinds are often posited as being the proper objects of scientific classification and inquiry insofar as the entities which natural kinds purport to pick out are those which share these important properties, which subsequently allow for fundamental scientific practices such as induction, explanation, and discovery.

Generally, essentialist accounts of natural kinds involve four main principles: (1) All and only the members of a kind share a common essence; (2): A kind's essence is a property or set of properties that all member of the kind must exhibit; (3): The properties that comprise a kind's essence are intrinsic, and; (4): A kind's essence is what causes the other properties associated with that kind.[15]

[14] i.e., *non-accidental*, scientifically relevant properties.

[15] The common example evoked in essentialism is the atomic structure of gold: the atomic structure of gold occurs in all and only pieces of gold. Gold's atomic structure must be possessed by all pieces of gold (non-accidental like being valuable to humans), and the atomic structure of gold causes pieces of gold to have the properties associated with that kind (e.g., conducting heat and electricity).

A crucial function of essentialist thinking is that it provides individuals with a foundation for categorization. However, when faced with the task of classifying biological and mental phenomena, traditional essentialism has been largely deemed ill-suited.[16] Nevertheless, there's reason to think that the way in which we tend to think about atypical members of essential kinds is intimately connected to the doxastic status debate's verbal dispute.

According to Newman and Knobe (2019), when individuals perceive a category as having a *causal essence,*[17] they tend to conceive of that category as having sharp boundaries:

> [A] person might think: "Different animals might look like tigers to different degrees, but ultimately, any animal is either a tiger or it isn't and that's all there is to it." (Newman & Knobe, p.12).

However, Newman and Knobe go on to describe two ways in which some essentialist categories are perceived as not having sharp boundaries. First, in cases *of graded membership*, a particular concept is associated with non-dichotomous criteria. On this account, some objects are taken as straightforwardly falling within or outside of a category, whereas other entities are perceived to have an intermediate status (p.12). To elucidate the intermediate status involved in graded membership, Newman and Knobe consider the concept TALL: some people are clearly tall, some are clearly not, and some individuals are taken to have an intermediate status such that it's not clear whether they should be characterized as tall or not (p.12).

---

[16] Kendler et al. (2011) argue that essentialism is ill-suited for classifying biological phenomena for the following two reason. First, the concept of an "essence" is not useful for describing developing and evolving organisms because these kinds of entities can significantly vary from another (social organizations, molecular interactions, etc.), even though they are of the same (natural) kind (p.1144). Second, essentialism assumes a single causal agent (p.1144). Yet, only some biological phenomena have been shown to result from single causal factors. Instead, the current evidence suggests that a large majority of diseases arise from a wide range of factors (e.g., genetic, metabolic, environmental).

[17] Causal essentialism refers to when the essence of a natural kind is taken to be what causes the superficial features of the kind.

Alternatively, in cases of *dual character*, individuals associate a particular concept with two distinct criteria: a set of superficial features, and a hidden essence (p.13). Here, individuals perceive a given object as being a member of a particular category if it fulfils both criteria, to not be a member of the category if it fails to fulfill either criteria, and objects that fulfill one criterion, but not the other are said to have a more confusing status (p.13). In cases of dual character, individuals effectively associate essentialized concepts with two distinct representations: a set of superficial features and a hidden essence (p.15).

Importantly, the "confusing status" associated with dual character is distinct from the intermediate status involved in cases of graded membership:

> It is not that there is a set of criteria that an object fulfills to an intermediate degree. Rather, it is that there are *two different* criteria. Thus, the very same object may be clearly a member of the category on one set of criteria but, at the same time, clearly not a member of the other (p.13).

In other words, in the case of dual character, the confusing status results from there being an object which fulfills two different criteria, whereas in the case of graded membership an object fulfills a set of criteria to an intermediate degree.

Evidently, scholars in the debate don't perceive the concept "delusion" as being circumscribed by sharp boundaries. For, we find that buried throughout the literature, scholars on both sides of the debate grant variation in cognitive attitudes *across individuals* with delusions.[18] However, the literature also suggests that scholars in the debate implicitly reject the possibility of delusions as having graded membership. For although doxastic and non-doxastic accounts are

[18] Scholars on the doxastic side of the debate frequently include a passing mention that *not all* delusions are doxastic states; likewise, non-doxastic scholars frequently mention that *not all* delusions are non-doxastic states. See Dub 2017, p.18; Currie 2000, p. 175-6, and Bortolotti 2009. However, I set aside this particular type of variation and will instead focus on what one might call the "intrasubjective-instability" of delusions, which refers to how within any *individual* case, one's delusion can exhibit an *odd* and *unstable* combination of features. Throughout §4, the intrasubjective-instability is discussed in greater detail.

both *prima facie* attractive positions,[19] they're taken by the parties in the debate to be incompatible positions. Hence, rather than positing a compatibilist position in accordance with the notion of graded membership, the dominant trend in the literature is to reject one or the other: delusions are either beliefs or not beliefs, there's no in-between.[20]

Yet, it's doubtful that matters would fare any better if scholars were to insist on categorizing delusions as having dual character. For as we have seen with Miyazono's account, aside from the superficial criteria (i.e., the features of delusions), it's not clear as to what the hidden criteria of delusions might be. It may be the etiological and causal source of delusions. But again, how do we distinguish the etiology of "deluded-beliefs" from say, the etiology of "deluded-imaginings"? Moreover, even if there does exist some significant hidden criteria, this position doesn't seem to be the one that's on many theorists' radar given that very few theorists elude to this position.

If delusions are taken as lacking sharp boundaries, yet scholars in the debate avoid appealing to the notion of graded membership or dual character, why do they continue to insist on characterizing delusions with pre-theoretical attitude terms? I believe there's reason to think that the answer to this question has to do with the way in which the combinations of features often exhibited in cases of delusions come across as being an unlikely pairing of features.[21] Rosch et al. write,

---

[19] On one hand, there are important respects in which the features of delusions often appear characteristic of belief (e.g., behaving as if the content of their delusion were true, involuntariness). On the other hand, the features of delusions may also appear patently uncharacteristic of belief (e.g., cases of *double-bookkeeping*, unresponsiveness to evidence).

[20] According to Egan (2009) delusions are an in-between state of imagination and belief – what he calls "bi-imagination." The motivation behind Egan's position is that it doesn't seem as though delusional individuals "straightforwardly" believe the contents of their delusions, but nor do they "straightforwardly" imagine the contents of their delusions (Egan, p.263). While I'm sympathetic to Egan's position, his account is rather vague and suggests a single set of features as being characteristic of delusions, which I will later argue against.

[21] e.g., unresponsiveness to evidence and involuntariness.

> The world is structured because real-world attributes do not occur independently of each other. Creatures with feathers are more likely to also have wings than creatures with fur, and objects with the visual appearance of chairs are more likely to have functional sit-on-ableness than objects with the appearance of cats. That is, combinations of attributes of real objects do not occur uniformly. Some pairs, triples, or ntuples are quite probable, appearing in combination sometimes with one, sometimes another attribute; others are rare; others logically or empirically do not occur (Rosch, et. p.383).

To frame the matter in line with the above passage, the combinations of features often presented in delusions seem quite improbable, and in order to get around this improbable and peculiar nature, theorists have attempted to logically explain away the oddness by characterizing delusions as atypical subsets of pre-theoretical attitudes, such as "beliefs" and "imaginings." However, doing so has pushed the debate into a verbal dispute, precisely because delusions are significantly unlike the types of attitudes which can be accounted for by pre-theoretical attitude concepts. Accordingly, if we wish to advance the doxastic status debate into more fruitful territory, we'll need to modify and expand the way in which we think about attitudes as natural kinds.

4 ADVANCING THE DOXASTIC STATUS DEBATE

In this section, I attempt to advance the doxastic status debate into more substantive territory by putting forward a new descriptive model for characterizing delusions, which I refer to as the *hetero-dynamic property cluster* model (HDPC). Fundamental to the HDPC model is the odd and unstable nature exhibited by HDPC kinds such as delusions. In order to exemplify this odd and unstable nature, in §4.1 I offer an examination and comparison of how two individuals with Capgras treat their delusions.

4.1 Two Cases of Capgras

Case 1 – "Fred":

"Fred" began to claim that his wife, "Wilma," had been replaced by an identical double (Lucchelli and Spinnler, 2007). On one occasion Fred went out to actively search the streets for his "real" wife. Upon returning home, Fred looked relieved to find his actual wife there and anxiously asked where she had been and why she had not told him she was going out. On a different occasion, Fred even urged the Wilma impostor to go with him to report the real Wilma's disappearance (p.189).

Lucchelli and Spinnler further note a time when Fred insisted that his wife had been replaced by an impostor yet failed to act as if he were bothered by his wife's disappearance. During this time, Fred even went so far as to treat the Wilma impostor in a kind and flirtatious manner. There were also times in which Fred's delusion appeared to be "inactive" in that during such times, Fred neither claimed that his wife was an impostor, nor did he behave as if she were.

Case 2 – DS:

DS, a 30-year-old man had a *tendency* to regard his parents as impostors (Hirstein and Ramachandran, 1997). I emphasize the word tendency here because like Fred, DS had swaying convictions about whether his parents were impostors. However, unlike with Fred, aside from verbal behavior (i.e., asserting that his father was an impostor), Hirstein and Ramachandran make no mention of DS acting in ways that we would likely associate as being characteristic of belief. That said, DS's verbal behavior appears to congruous with the odd and unstable nature of delusions.

On one occasion, DS was asked why the impostor was pretending to be his father. To this, DS responded by saying, "That is what is so surprising, Doctor -- why should anyone want to pretend to be my father? Maybe my father employed him to take care of me – paid him some money so that he could pay my bills…" (p.438). DS's response to this question appears to be

indicative of an attempt to search for auxiliary hypotheses that would make the delusional content *cohere* with the rest of his belief system. DS's father attempted to exploit the fact that his son was attempting to make the delusional content cohere with the rest of his beliefs by "tricking" his son out of his delusion: DS's father walked into his son's room and announced "The man you have been with all these days is an impostor – he isn't really your father. I have sent him away to China. I am your real father – it's so good to see you son" (p.439). What's relevant for the following discussion is not only that this trick worked on DS, but that it only worked briefly. DS reverted to his original delusion a week later.

4.2 The Odd and Unstable Nature of Delusions

In §3.1, it was stated that both doxastic and non-doxastic theorists largely agree on the following features as being characteristic of delusions: (i) involuntariness; (ii) unresponsiveness to evidence and insensitivity to evidence: (iii) behavioral circumscription, and; (iv) lack of coherence. Assessing the features of Fred's and DS's delusions, we can certainly see how each individual can be taken to exhibit (at least some of) these features. For instance: Although DS appears to exhibit some responsiveness to evidence insofar as he revises his attitude in light of his father's "trick," his responsiveness to evidence is only temporary, perhaps alluding to the involuntary nature of his delusion. Likewise, whether or not Fred perceives Wilma to be the "actual" Wilma appears to be involuntary as well. Additionally, DS's behavioral circumscription is evident given that he only makes verbal assertions pertaining to the content of his delusion and performs hardly any behaviors that would be relevant to the content of his delusion (e.g., calling the police, searching for his actual father). In Fred's case, his behavioral circumscription is captured during the times in which he asserts that Wilma is an impostor, but acts otherwise unphased by this (e.g., treating Wilma in a flirtatious manner).

However, an issue with taking (i)-(iv) to be the characteristic set of features exhibited by delusions can be captured by what I refer to as the "intrasubjective-instability" of delusions, which refers to how within any *individual* case, one's delusion can exhibit an odd and *unstable* combination of features. The unstable nature of delusions, which is captured by the way in which the features of one's delusion can change quickly and without any obvious reason as to why, is evident in both Fred's and DS's cases. For instance, in Fred's case, whether or not he takes Wilma to be the "actual" Wilma seems to change unpredictably. So too, whether Fred exhibits behavioral circumscription seems change unpredictably (e.g., actively searching for his real wife versus asserting that his wife is an impostor yet treating the impostor flirtatiously). Likewise, in DS's case, we see wavering assertions of his delusion, and the individual features of his delusion appear to waver as well (e.g., temporary responsiveness to evidence). Crucially, because delusions are unstable (i.e., the features of one's delusion can quickly waver), to say that (i)-(iv) are the characteristic set of features would be misleading, since at any one time a delusional individual *might* exhibit (i)-(iv), but they might also exhibit very few of these features, and perhaps even features that come across as antithetical to (i)-(iv). This last point brings us to the notion of delusional features as being *odd.*

The oddness of delusional features refers to how, unlike with beliefs and imaginings, the features of delusions do not appear to reinforce one another, and in some cases, the features of one's delusion may even appear to undermine one another. For example, revisiting Elaine's scenario (see §2.1), her involuntary belief that her roommate isn't home reinforces her ability to make rational behavioral choices. For instance, waiting for her roommate to return in order to talk with her. Conversely, in Fred's case, the involuntary delusional content that his wife is an impostor doesn't provide a similar reinforcement. For like Elaine, Fred is saddled with some

involuntary content that he cannot choose to ignore, yet unlike Elaine, if Fred treats the content of their delusion as reflecting reality, he'll encounter practical difficulties if he chooses to let the delusion guide his behavior in a belief-like manner. For instance, Fred will not find his "actual" wife by searching for her.

The intrasubjective-instability of delusions is relevant to the doxastic status debate in that it demonstrates how the odd and unstable nature of delusions cannot be adequately accounted for if we are to insist on characterizing delusions with pre-theoretical attitude terms, such as "beliefs" and "imaginings." In particular, what makes the use of such pre-theoretical attitude terms unfit for capturing the odd and unstable nature of delusions is two-fold. First, the two attitudes which delusions often bear resemblance to are one's whose features appear to be highly incongruous with one another (i.e., beliefs and imaginings; see §2.2). Thus, to characterize delusions as either "beliefs" or "imaginings" comes across as misleading since the features often exhibited in delusions aren't mere *slight* deviations from prototypical beliefs or imaginings, but often a mixture of directly opposed features of these attitudes. Second, because delusions are unstable, attaching a single label onto such an attitude likely won't be productive since at any given moment the features of one's delusion can change significantly (e.g., closer to or further from what we expect from prototypical beliefs). Below, Table 2 highlights the intrasubjective-instability of delusions.[22]

Table 2: Attitude Features: Belief, Imagining, and Delusion

(Attitude)	**Belief**	**Imagining**	**Fred's Delusion**	**DS's Delusion**	**Delusions (general)**
(Feature)					

[22] Note that the feature of context dependence is unknown in the case of delusions. This particular feature will be discussed in greater detail the final section of this paper. Also note that the last column is based on what both sides of the debate have said regarding the features of delusions.

Responsiveness to Evidence	**Yes**	**No**	**Mostly No**	**Mostly No**	**Mostly No**
Voluntary	**No**	**Yes**	**No**	**No**	**No**
Context Independence	**Yes**	**No**	**?**	**?**	**?**
Behavioral Guidance	**Yes**	**Yes (limited)**	**Yes & No**	**Mostly No**	**Yes & No**
Coherence	**Yes**	**Yes**	**Mostly No**	**Mostly No**	**Yes & No**

To address the intrasubjective-instability of delusions, in §4.4, I put forward the hetero-dynamic property cluster (HDPC) model. According to this descriptive model, delusions are mental states characterized by an odd and unstable cluster of features, which when manifested, contributes to delusion's failure to resolve neatly into any one kind of pre-theoretical cognitive attitude. Prior to presenting the HDPC model, in §4.3, I discuss Richard Boyd's (1991) *homeostatic property cluster* (HPC) model of natural kinds.

4.3 Homeostatic Property Clusters

Notably, the legitimacy of mental health disorders as proper targets of scientific inquiry has been a contentious issue (see Szasz 1960, Zachar 2001, Beebee & Sabbarton-Leary 2010), which can largely be traced back to the "anti-psychiatry" movement in the 1960s. The father of the anti-psychiatry movement, Thomas Szasz, argued that the concept *mental health disorder* not only fails to pick out a "real" kind in nature, but that it's nothing more but a "convenient myth" whose function is to disguise and "render more palatable the bitter pill of moral conflicts in human relations" (Szasz, 1960, p.117). Essentially, Szasz's claim that mental health disorders are not "real" is to say that what we typically refer to as "mental illnesses" (e.g., delusions, depression, anxiety disorders) are significantly different from diseases of the brain because they fail to correspond to objective abnormal brain anatomy in the same way that for instance,

epilepsy does (Szasz, p.115).[23] The upshot of the anti-psychiatry movement then, is that mental disorders can only exist as a type of folk concept, and hence cannot be proper targets of scientific (empirical) classification and investigation.

Several philosophers have challenged Szasz's position by arguing that psychiatry's objects (i.e., mental health disorders) are proper targets of scientific classification and inquiry because they *are* natural kinds (see Tekin 2016, Kincaid & Sullivan 2014, Samuels 2009). However, given that essentialist models of natural kinds have largely been rejected for classifying mental phenomena, several philosophers have considered alternative models of natural kinds for classifying mental phenomena.

As far as I can tell, Richard Samuels (2009) is the first to explicitly discuss delusions in connection with Richard Boyd's (1991) *homeostatic property cluster* (HPC) model of natural kinds.[24] Although I'm sympathetic to Samuels's position, as will be discussed, the HPC model appears to face significant challenge for characterizing delusions. Nevertheless, the HPC model serves as a useful framework for the hetero-dynamic property cluster model (described in §4.4). Accordingly, prior to discussing the hetero-dynamic property cluster model, I present a basic overview of Richard Boyd's homeostatic property clusters model.

Unlike traditional accounts of natural kinds (e.g., essentialism), kind membership on the HPC model is flexible and allows for an imperfect clustering of properties. Crucially, what matters on the HPC model is that the similarities among members of a kind are *stable* enough to yield reliable (even if imperfect) inferences and predictions about members of the kind.[25]

---

[23] Caused by abnormal electrical activity in the brain and can be observed with imagining tests.

[24] Samuels endorses Boyd's *homeostatic property cluster* (HPC) theory of natural kinds to defend the position that delusions are natural (HPC) kinds.

[25] Samuels helpfully explains that the properties of a natural kind are not *logically* related but contingently and non-accidentally associated. Samuels writes, "Consider a paradigmatic example of a natural kind: water. Samples of water tend to possess a wide array of characteristics – transparency, potability, specifiable boiling and freezing

Moreover, on the HPC model, natural kinds are groups of entities that exhibit a contingently clustering family of properties, held together by underlying homeostatic causal mechanisms.[26] These homeostatic mechanisms undergird the inductive power of HPC kinds by ensuring that members of kind will be alike on the basis that one property will often co-occur with other particular properties.

As was hinted at in §2.1-2.2, there's reason to maintain that beliefs and imaginings conform to the HPC model of natural kinds insofar as they tend to exhibit distinct clusters of properties which often co-occur with other properties due to a reinforcement among the kind's features.[27] To characterize beliefs and imaginings as HPC attitudes isn't to claim that all instances of these attitudes (e.g., all token beliefs) will exhibit an identical set of features. For such a claim would conflict with an essential characteristic of HPC kinds, namely, that kind membership on the HPC model doesn't impose strict necessary conditions. As stated above, what kind membership on the HPC model requires is for the similarities among members of a kind to be *stable enough* to yield reliable inferences and predictions about members of a kind. Accordingly, a mental state needn't exhibit *every* feature of its exemplary kind, but it must exhibit enough features to make it a relatively stable kind.

Furthermore, maintaining that beliefs and imaginings are HPC kinds doesn't imply that the attitude features nor the attitude itself are fixed. For instance, it's possible for one to hold a belief that's evidentially based early on, but as time passes its degree of responsiveness to

---

points, and so on. Moreover, these characteristics are not logically (or conceptually) necessary properties of water samples in the way that, say, being unmarried is a necessary property of bachelors" (p.52).

[26] Kendler, Zachar, and Craver's (2011), offer a "modified" version of Boyd's HPC model, viz., the *mechanistic property cluster* (MPC) model of kinds. However, it's not entirely clear whether there's a substantial difference between HPC and MPC kinds. It seems that the term "mechanistic" is favored primarily because the term "homeostatic" is misleading outside of biology (p.1146).

[27] Additionally, "non-standard" attitudes (see Van Leeuwen 2014 on religious credences) may also be such that they conform to the HPC model of kinds.

evidence decreases. As long as the attitude in question still has enough features of belief to make it a stable kind, then it would still be appropriate to characterize the state as a belief on the HPC model. That said, although the HPC model allows for more flexibility than traditional essentialist models, there's still only so much flexibility: if too many features of a mental state that initially looked like a belief end up diverging from the features typically associated with belief, then there's reason to no longer characterize the state in question as a belief. In such a scenario, a change in attitude characterization would be warranted since the property cluster of the state's proposed kind has now deviated too much from its homeostatic prototype.

At first blush, the HPC model looks promising for characterizing delusions. Afterall, the flexibility of kind membership on the HPC model seems fit for capturing the way in which delusions diverge from prototypical beliefs and imaginings. But herein likes the challenge: although the flexibility of kind membership on the HPC model has typically been advantageous for categorizing biological and psychological kinds, when faced with the task of characterizing delusions, the HPC model's flexibility makes it relatively easy for scholars to trivialize the *oddness* of the features of delusions: doxastic theorists can grant that delusions have an odd set of features while simultaneously maintaining that they're *not odd enough* to be rejected as beliefs; likewise, non-doxastic theorists can grant that delusions have an odd set of features and explain their oddness by positing a non-doxastic attitude to characterize them.

Moreover, on the HPC model, if the property cluster of a proposed kind deviates too much from the homeostatic prototype, then it will no longer be a *stable* kind and hence its status as an HPC kind is threatened. Given what we've seen with Fred and DS, there's reason to believe that delusions are *not* stable kinds in the way that HPC attitudes such as beliefs and imaginings tend to be. In fact, if the HPC model were apt for characterizing delusions, then

identifying the correct attitude involved in delusions should allow us to make reliable inductive generalizations about delusions on the basis of their categorization. Consider the following two mental states:

(a) Lucy *imagines* that her mother is an impostor.

(b) Lucy *believes* that her mother is an impostor.

Because paradigmatic beliefs and imaginings tend to conform to the HPC model, it's reasonable to infer that if Lucy merely *imagines* that her mother is an imposter she won't actually act as if it were true that her mother has been replaced by an impostor (e.g., Lucy won't call the police or search for her actual mother). Conversely, if Lucy *believes* that her mother is an imposter, we'd expect her to take these types of actions. Yet, characterizing one's mental state as a delusion has limited utility, especially when compared to characterizing one's mental state as a belief or an imagining.

To be fair, even without a unanimously agreed upon definition of "delusion," mental health care professionals as well as laypeople can typically agree on when something should be characterized as delusional.[28] Yet unlike cases of standard cognitive attitudes, because of their unstable nature, labeling something as a "delusion" doesn't allow us to make reliable generalizations or predictions. For instance, one might assume that we should be able to reliably infer that delusions won't respond to counterevidence since, like imaginings, delusions tend to lack evidential vulnerability. But in the case of delusions, such inferences are extremely unreliable.[29]

---

[28] For folk-epistemology on delusions see Murphy (2012).

[29] For instance, DS *did* temporarily respond to counterevidence when his father implemented his "trick." Similarly, characterizing one's mental state as a belief or an imagining should yield reliable inferences regarding behavioral guidance. But as we have seen with Fred, delusional behavior can waver between belief-like behavior and imagining-like behavior for no obvious reason. One might object that DS's temporary responsiveness to evidence is consistent with HPC attitudes since the HPC model allows for an imperfect clustering of properties. But unlike HPC attitudes whose features tend to either remain stable or require sufficient reason to change, DS's attitude revision

However, failure to correspond to essentialist or HPC models of natural kinds doesn't mean that the attempt to characterize delusions will inevitably lead down an explanatorily useful dead-end. In what follows, I turn to my attempt to advance the doxastic status debate into more substantive territory by adapting Boyd's (1991) HPCs model into the *hetero-dynamic property clusters* (HDPCs) model.

4.4 Hetero-Dynamic Property Clusters

According to the hetero-dynamic property cluster (HDPC) model, delusions are mental states characterized by an odd and unstable cluster of features. When manifested, this odd and unstable nature gives delusions the appearance of straddling the line between two distinct types of attitudes. To some extent then, delusion's odd and unstable nature has served as a catalysis to the doxastic status debate in that it inexorably contributes to delusion's failure to resolve neatly into any one kind of pre-theoretical cognitive attitude. I will argue, however, that by adopting the HDPC model, we can begin to make sense of the peculiar -- odd and unstable -- nature of delusions, while still maintaining a degree of explanatory potential comparable to that of essentialist and HPCs models of natural kinds. I begin the discussion of hetero-dynamic property clusters by describing how they diverge from homeostatic property clusters.

The fundamental distinction between HPCs and HDPCs concerns the stability of each respective kind. Specifically, while HPC kinds are stable kinds whose stability result from a mutual reinforcement among features, the features of HDPC kinds often fail to reinforce one another in a straightforward manner and may even appear incompatible with one another. To say that HDPCs are *odd* refers to this notion of HDPC kinds as often displaying an inconsonant

---

was neither stable, nor were there any obvious reasons for its change. So, while it's true that an HPC mental state doesn't require a necessary set of features, delusional mental states still appear to differ from HPCs in that their features quickly waver, and they do so for reasons that are currently unclear.

combination of features. Ultimately, the oddness of HDPCs is the source of their instability: because HDPC features don't reinforcing another, there's nothing to hold them in place, so to speak. Hence, HDPCs are *unstable.* As a result of this instability, the features of HDPCs can change quickly and without any obvious reason as to why. With respect to attitudes, the unstable nature of HDPCs essentially dampens restrictions on the mental state's movement both within the property space of a particular attitude and among one property space to another. When manifested, such instability allows the features of an HDPC attitude to freely wavier between two distinct property clusters. To better clarify the differences in stability between HPC and HDPC attitudes, consider Figures 1 and 2.

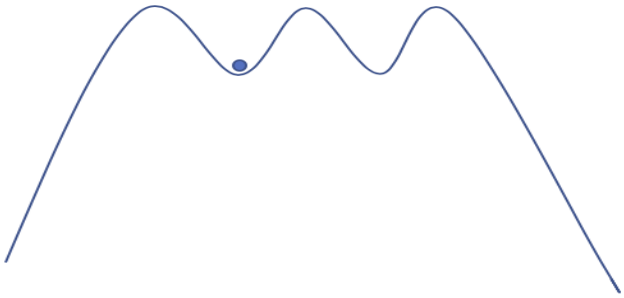

Figure 1

Let the (metal) ball in Figure 1 represent an HPC attitude and the valley where the ball resides represent the designated property space of belief. When the ball is located at the lowest point of the belief-valley, the corresponding mental state exhibits all of the features typically associated with belief (see §2.1). The greater the ball's distance from the belief-valley's lowest point, the fewer features of belief exhibited by the mental state. Additionally, let nearby valleys represent the property spaces of other attitudes. In this case, nearby valleys should be taken to be ones which have some features in common with prototypical beliefs. For instance, the property

spaces of an acceptance or an assumption would likely be nearby (see Dub, 2017; Bratman, 1992 for more on acceptances).

Suppose that the individual who holds the belief, stands at the belief-valley's lowest point and gives the ball a single push. This push will cause the ball to move from its initial location, however the effects of gravity will eventually cause the ball to come to rest at its original spot *or close to it*. The fact that the ball needn't finish its trajectory at its *exact* starting position corresponds to the flexibility of HPC kind membership. That is, as long as the ball remains located somewhere within the valley (i.e., the property space of belief), then despite any slight deviations from its original starting point, the corresponding mental state remains characterized as a belief. Furthermore, gravity's effect on the ball's trajectory corresponds to the underlying homeostatic mechanisms of HPC kinds. In the case of HPC attitudes, the functions of homeostatic mechanisms are comparable to a type of epistemological force which, similar to gravity, pulls the attitude's features toward one another in order to form a distinct cluster of features.

Of course, the effects of gravity don't render it impossible for the individual to move the ball up the valley's peak and into a nearby valley. Likewise, the epistemological force of HPCs neither prohibits changes in the attitude's features nor changes from one attitude to another. However, moving the ball from one valley into another is more difficult than just nudging the ball with a single push. If we suppose that the individual in the scenario has several tools available to them for moving the ball, we can see how moving the ball from one valley to another is an achievable process. Similarly, one's transition from an acceptance to a belief is a gradual process that often requires well-functioning cognitive tools.

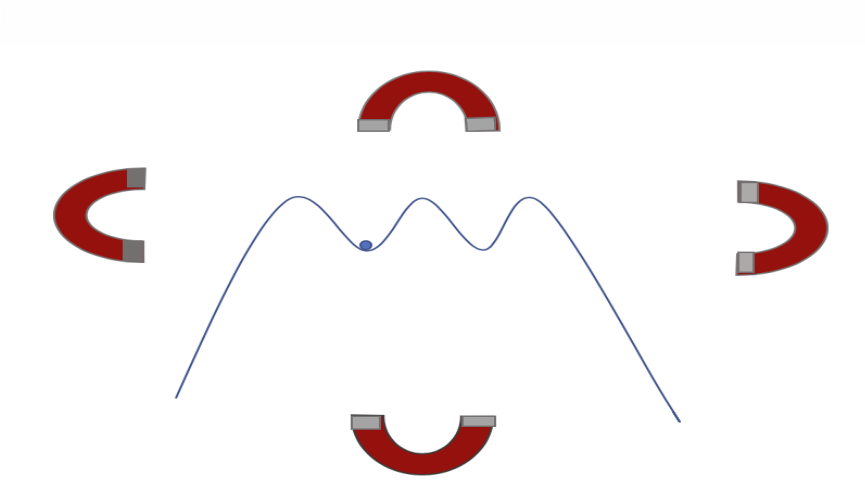

Figure 2

Let the (metal) ball in Figure 2 represent an HDPC attitude and the left and right valleys represent the property space of belief and imagining, respectively. While it's still the case that nearby valleys should represent the property spaces of attitudes that share similar features, we can conceive of Figure 2 as a sort of condensed attitude map which excludes all properties spaces that don't correspond to beliefs or imaginings. As we will come to see, the distance between the "belief-valley" and "imagining-valley" in the HDPC case is not as significant as it is in the HPC case. Furthermore, unlike with Figure 1, the peaks and valleys in this scenario are surrounded by magnets.

Again, suppose that there's someone standing at one of the valley's lowest points and that they also have tools available for moving the ball. Unfortunately for this individual, some of the tools are damaged or malfunctioning. Unsurprisingly, the condition of the tools presents a challenge to the individual's ability to move the ball. Worse yet, there also exists an unknown mechanism which causes the magnets to unpredictably shift locations (e.g., farther or closer to one another). Given the circumstances, the individual's ability to move the ball greatly depends

on the orientations of the magnets. If the individual wishes to move the ball, the best-case scenario would be for the magnets to be positioned far enough away from one another so that they neither attract nor repel the metal ball. In this case, the primary challenge for the individual is their ability to make do with the malfunctioning tools. The worse-case scenario, however, would be for the magnets to be oriented in such a way that would ultimately cause the individual to have no control over the ball's movement.

If we take the underlying causal mechanism(s) of delusions to play a similar role as the magnets in Figure 2 (rather than the role of gravity in Figure 1), the instability of delusions (i.e., how the features of delusions can quickly change) becomes a bit easier to make sense of: In the HPC scenario the movement of the ball is gradual, consisting of its being pushed up the peak of a property space. Conversely, in the HDPC case (Figure 2), the movement of the ball needn't be a gradual one. If the magnets shift abruptly, then movement of the ball will also be quite sudden. Moreover, because the ball's movement in the HPC case is a gradual one consisting of movement up a particular property space's peak and into nearby valleys which have features in common with the initial property space, the transition from one HPC property space to another is often inconspicuous. Hence, a mental state located half-way between the property space of belief and the property space of an assumption, likely won't strike us as peculiar.[30] Conversely, due to the possible movements of the magnets in the HDPC case, the process of moving the ball from one valley to another needn't take the form of a gradual movement up the property space's peak. In this scenario, the magnets are capable of *hovering* the ball in positions that would be difficult to achieve in the HPC case. These irregular positions can be such that they correspond to property spaces whose features do not share many features (e.g., beliefs and imaginings). One

[30] However, we may find such a state odd if it persists for too long. In between states of HPC attitudes may be related to cases of weakness of will, or implicit attitudes that we don't take ourselves to have.

way to think about these irregular positions would be to think about the ball's movement in the HPC case as being restricted to two-dimensional movement, such as on an X, Y axis, whereas the movement of the ball in the HDPC case includes positions of depth (X,Y,Z axis).

A worry that might arise is whether the HDPC model can achieve the fundamental scientific practices often associated with natural kinds. Although I consider this to be a reasonable worry, I think it's quite plausible that the HDPC model can allow us to produce inductive generalizations, explanations, and so forth. However, due to the instability of HDPC kinds, it's likely that most generalizations and predictions about HDPC kinds will be less detailed than those of stable kinds. For unlike with HPCs wherein kinds are stable point (similar to valleys) which allow us to generalize about future states based on the kind's underlying mechanisms (similar to gravity's effects on the ball), with HDPCs, without fully understanding the mechanisms which cause the magnets to behave in the way that they do, we won't be able to predict where the ball will ultimately end up, regardless of where the ball initially resides in Figure 2. That said, as we come to better understand the relevant mechanisms of HDPC kinds, we'll be in a position to make predictions about *possible* movements. However, since HDPCs are unstable, the philosopher's and psychiatrist's understanding of the nature of delusions and will likely end up having more in common with meteorologist's understanding of weather and the astronomer's understanding of planetary orbit than with the biologist's understanding of cells. Of course, the weather forecast is nothing to sneeze at: I always bring my umbrella when it calls for rain.

5 CONCLUSION

In order to better understand the nature of delusions, scholars have sought to determine which cognitive attitude best characterizes delusions. From this endeavor, the doxastic status

debate emerged. Despite efforts, there has been little progress in settling the doxastic status debate. In this paper, I have argued that the debate has been stymied because it's largely a verbal dispute and that attempts to characterize delusions with pre-theoretical attitude terms, such as "beliefs" and "imaginings" has greatly contributed to the verbal dispute.

Additionally, I've argued that attempting to characterize delusions with pre-theoretical attitude terms overlooks an essential quality of delusions. Namely, the intrasubjective-instability of delusion, wherein within any *individual* case, one's delusion can exhibit an *odd* and *unstable* combination of features. To address the intrasubjective-instability of delusions, I've put forward the hetero-dynamic property cluster (HDPC) model. According to this descriptive model, delusions are mental states characterized by an odd and unstable cluster of features, which when manifested, contributes to delusion's failure to resolve neatly into any one kind of pre-theoretical cognitive attitude. I've also suggested that by adopting the HDPC model, the peculiar (odd and unstable) nature of delusions can be accounted for while still maintaining a degree of explanatory potential comparable to that of essentialist and HPCs models of natural kinds. However, there remains an important issue which has yet to be addressed in this paper. Specifically, what are the causal mechanisms involved in delusions that underpin their odd and unstable nature? Of course, the HDPC model is only a descriptive model and discovering the correct causal mechanisms involved in delusions is a matter of future empirical work. That said, I conclude this paper with a few comments on how context dependence/ independence and the default mode network may point towards a possible causal mechanism which gives rise to the HDPC nature of delusions.

It should go without saying that determining the features of delusions is deemed as an important endeavor to our understanding of delusions. Yet one feature in particular that hasn't received much attention in the doxastic status debate, but which may be invaluable to our

understanding of delusions is context dependence/independence. Perhaps the reason why context dependence/independence hasn't received much attention in the doxastic status debate is because it's not clear how to assess this feature in the case of delusions. Fittingly, Table 2 presented in §4.2 fails to specify whether delusions are context independent or context dependent. However, in what follows, I attempt to explore the role of context dependence/independence in cases of delusions.

In §2.1, context independence[31] was briefly explained by stating that if one believes *p*, *p* is taken as a given in practical reasoning and decisions, regardless of the practical setting one is in; where practical setting is "the dimension of a cognitive attitude (independent or dependent) that specifies the sorts of situations in which it guides behavior" (Van Leeuwen, 2014, p.702). Both Bratman (1992) and Van Leeuwen (2014) maintain that belief is the *only* cognitive attitude which is practical setting *independent.*[32] Roughly, the distinction between context dependent and context independent attitudes is that an in the latter case, the cognitive attitude will guide one's behavior in all settings in which the content of the attitude is relevant, whereas an attitude that's practical setting dependent will only guide actions in a particular setting or context. Van Leeuwen provides the following example to elucidate this distinction:

> Adult experimenters pretended to be on a picnic with child subjects and took an *actual* bite out of a piece of Playdough they were pretending was a cookie. If children confuse reality with pretense, as many think, the bite would have been no surprise, since the subjects would have taken the adult to have bitten *a cookie.* But the children were surprised – even disconcerted – which showed that they were tracking the piece of Playdough all along, *even in the context of make-believe.* Call this *continual reality tracking* (Van Leeuwen, p.701, discussion of a study by Golomb and Kuersten, 1996).

[31] What Bratman refers to as *context independence*, Van Leeuwen refers to as *practical setting independence.* These terms are essentially equivalent, and as such, can be used interchangeably.

[32] Where practical setting independence is defined by Van Leeuwen as, "A cognitive attitude *x* is practical setting independent if and only if *x* guides behavior in all practical settings in which *x*'s content is relevant to the agent's behaviors (Van Leeuwen, p.702), and Bratman describes context independent as: At any one time a reasonable agent normally either believes something (to degree *n*) or does not believe it (to that degree). She does not at the same time believe that *p* relative to one context but not relative to another (Bratman, p.3).

Although the term double-bookkeeping commonly refers to *delusional* double-bookkeeping (wherein delusional individuals appear to perceive both reality and their delusion as being real yet remain largely unaffected by the discrepancies between the two "realities," see §3.1), there's a sense in which individuals without delusions engage in double-bookkeeping in everyday life. For instance, in the above passage from Golomb and Kuersten we can say that the children engage in double-bookkeeping when they participate in make-believe play in that they're reflexively able to partition their representations of the world between make-believe and reality. This is all to say that for individuals without delusions, it doesn't seem difficult for one to determine which "relevant practical setting" they're in. In fact, it seems that knowing which practical setting one is in is often an unconscious recognition. In the case of delusional double-bookkeeping, however, it seems that individuals are unable to successfully partition their representations of delusion and reality (i.e., the individual's ability to reality track is malfunctioning). If an individual's ability to continually track reality is diminished, there's reason to think that they'll have difficulty accurately tracking which practical setting they're in (e.g., if I cannot track whether I imagine or believe that my mother is an impostor, the way I act on this representation will be affected).

Consider persecutory delusions, the delusion that people are out to get you, or harm is going to occur. What is the relevant practical setting here? Would it be any setting where the subject is around other people? Only when certain people or certain situations are present? It's not clear from a third person perspective, and most likely not from the first-person perspective, when the "relevant setting" for persecutory delusion has been "entered." Related to this notion of double-bookkeeping and reality tracking, a question worth asking is what individuates the

"books" that are being kept during instances of double-bookkeeping. In cases of delusions, individuals may have two sets of books, a "Reality book" and a "Delusion book." But unlike with books such as an "Imagining book" a "Religious credence book," and so forth, the delusional individual isn't always able to track which book is the Reality book and which book is the delusion book. Furthermore, given that delusional and non-delusional individuals can have more than one book up and running, what is it that allows one to differentiate books? Whatever it is that allows for one to differentiate books seems to be at least one of the mechanisms that's malfunctioning in cases of delusions.

Although identifying the mechanism related to reality tracking and that the ability to distinguish books from one another is a matter of future empirical work, there's already some empirical work which seems interestingly relevant: Gerrans (2013) has proposed that schizophrenic delusions are the result of a hyperactive default mode network (DMN) or a DMN whose activity isn't properly modulated by prefrontal mechanisms:

> [Decontextualization] refers to cognitive processes which use representational formats which are impersonal and context-free. Decontextualisation is not itself a specific form of inference, such as logical reasoning or hypothesis confirmation, reading maps or graphs, but a necessary cognitive precondition for such forms of context-independent cognition. […] What these tasks have in common is that they require the subject to represent information in a context-free way to generate a novel solution (Gerrans, p.90).

In a different article, Gerrans (2014) proposes that dream experience may be relevant to delusions of misidentification (e.g., Capgras) in that both dreams and delusions involve experiences that are highly inconsistent with background knowledge, and that it's often the case that in both dreams and delusions, background knowledge isn't used and experiences aren't

explained or explained away; instead, they're taken at face value.[33] According to Gerrans, most dream theorists explain issues of doxastic reality testing in terms of deactivation of neural circuity required for the necessary metacognitive processing, specifically, the deactivation of the (DMN).

Gerrans's discussion on the DMN and decontextualization appears to be consistent with the HDPC model of delusions, and the lack of balance between the operation of the DMN and decontextualized processing mechanisms might create the precise conditions for the kind of instability characteristic of delusions. To put it another way, one role of the prefrontal cortex's decontextualized processing is to enforce epistemic stability. For instance, if confronted with a representation that is less subject to voluntary control than a typical imagining, but is appreciably inconsistent with evidence, this mechanism enforces a placement of the representation as either a belief-like state whose truth requires explanation or an imaging-like state whose vivacity needs to be ignored. Accordingly, a potentially promising avenue for future research will be to explore the interaction between the DMN and the prefrontal mechanisms that are responsible for decontextualized processing.

[33] Gerrans suggests that this is connected to "doxastic reality testing" which refers to "one's ability to test experiential beliefs for consistency and correctness with background knowledge, thereby establishing whether they correspond to the world as it is, independent of the subject's mind.

REFERENCES

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5.* American Psychiatric Publications Incorporated.

Bayne, T., & Pacherie, E. (2005). "In Defense of the Doxastic Conception of Delusions" in *Mind & Language, 20*(2):163-188.

Beebee, H., & Sabbarton-Leary, N. (2010). "Are Psychiatric Kinds 'Real'? in *EuJAP, 6*(1):11-27.

Bennett, J. (1990). "Why is belief involuntary?" in *Analysis, 50*(2): 87-107.

Bortolotti, L. (2009). *Delusions and Other Irrational Beliefs.* Oxford University Press.

Bortolotti, L. (2011). "Double Bookkeeping in Delusions: Explaining the Gap between Saying and Doing" in *New Waves in Philosophy of Action.* Palgrave Macmillian, London. 237-256.

Boyd, R. (1991). "Realism, antifoundationalism, and the enthusiasm for natural kinds" in *Philosophical Studies,* 61:127-148.

Bratman, M. E. (1992). “Practical reasoning and acceptance in a context” in *Mind*, *101*(401), 1-15.

Chalmers, D. (2011). “Verbal Disputes” in *Philosophical Review, 120*(4): 515-566.

Coltheart, M. (2007). “Cognitive neuropsychiatry and delusional belief” in *The Quarterly Journal of Experimental Psychology, 60*(8):1041-1062.

Currie, G. (2000). “Imagination, Delusion and Hallucinations” in *Mind & Language, 15*(1): 168-183.

Currie, G., and Ravenscroft, I. (2002). *Recreative Minds: Imagination in Philosophy and Psychology.* Oxford University Press.

Dub, R. (2017). “Delusions, Acceptances, and Cognitive Feelings” in *Philosophy and Phenomenological Research,* 94:27-60.

Foster Hanson, E., Leslie, S., Rhodes, M. (2016). “How Does Generic Language Elicit Essentialist Beliefs?” in *CogSci.*

Gerrans, P. (2013) “Delusional Attitudes and Default Thinking” in *Mind & Language, 28*(1): 83-102.

Gerrans, P. (2012). “Dream experiences and a revisionist account of delusions of Misidentification” in *Consciousness and Cognition,* 21:217-227.

Ginet, C. (2001). “Deciding to Believe” in *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, 63-76.

Golomb, C., and Kuersten, R. (1996). “On the transition from pretense play to reality: What are the rules of the game?” in *British Journal of Developmental Psychology. 14,* 203-217.

Hirstein, W., and V., Ramachandran. (1997). “Capgras Syndrome: a novel probe for understanding the neural representation of the identity and familiarity of persons” in

Proceedings. Biol Sci, 264(1380): 437-444.

Kendler, K., Zachar, P., & Craver, C. (2011). "What kind of things are psychiatric disorder?" in *Psychological Medicine, 41*(1143-1150).

Kincaid, H. & Sullivan, J. (2014). *Classifying psychopathy: Mental kinds and natural kinds:* 1-10.

Lucchelli, F., and Spinnler, H. (2007). "The case of lost Wilma: a clinical report of Capgras Delusion" in *Neurological Science, 28*(4): 188-195.

Miyazono, K. (2019). *Delusions and Beliefs: A Philosophical Inquiry.* Routledge Focus on Philosophy.

Murphy, D. (2012). "The Folk Epistemology of Delusions" in *Neuroethics.* (5); 19-22.

Newman, G., & Knobe, J. (2019) "The Essence of Essentialism" in *Mind & Language.* 10.1111/mila.12226.

Nichols, S. (2004). "Imagining and Believing: The Promise of a Single Code" in *The Journal of Aesthetics And Art Criticism, 62*(2): 129-139.

Rose, D., Buckwalter, S., & Turri, J. (2014). "When Words Speak Louder Than Actions: Delusions, Belief, and the Power of Assertion" in *Australasian Journal of Philosophy, 92*(4): 1-18.

Rose, D., Machery, E., Stich, S., Alai, M., Angelucci, A., Berniūnas, R., Buchtel, E.E., Chatterjee, A., Cheon, H., Cho, I.-R., Cohnitz, D., Cova, F., Dranseika, V., Lagos, Á.E., Ghadakpour, L., Grinberg, M., Hannikainen, I., Hashimoto, T., Horowitz, A., Hristova, E., Jraissati, Y., Kadreva, V., Karasawa, K., Kim, H., Kim, Y., Lee, M., Mauro, C., Mizumoto, M., Moruzzi, S., Olivola, C.Y., Ornelas, J., Osimani, B., Romero, C., Rosas, A., Sangoi, M., Sereni, A., Songhorian, S., Sousa, P., Struchiner, N., Tripodi, V., Usui, N., del Mercado, A.V., Volpe, G., Vosgerichian, H.A., Zhang, X. and Zhu, J. (2017),

Behavioral Circumscription and the Folk Psychology of Belief: A Study in Ethno Mentalizing. Thought: A Journal of Philosophy, 6: 193-203.

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). "Basic Objects in Natural Categories" in *Cognitive Psychology,* 8:382-439.

Samuels, R. (2009). "Delusions as a Natural Kind" in *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives.* Oxford University Press, 49-79.

Sharma, A., Garuba, M., &, Egbert, M. (2009). "Capgras Syndrome in a Patient with Multiple Sclerosis: A Case Report" in *Physicians Postgraduate Press, Inc,* 11(5): 274.

Tekin, S. (2016). "Are Mental Disorders Natural Kinds?: A Plea for a New Approach to Intervention in Psychiatry" in *Philosophy, Psychiatry, & Psychology, 23*(2):147-163.

Van Leeuwen, N. (2011). "Imagination is where the action is" in *The Journal of Philosophy, 108*(2):55-77.

Van Leeuwen, N. (2014). "Religious credence is not factual belief" in *Cognition,* 133:698-715.